MEDWELL PUBLISHING
Journal of Animal and Veterinary Advances 8 (3): 481-487, 2009
ISSN: 1680-5593

Examining Structural Distribution of Livestock in Eastern and South-Eastern Anatolia of Turkey by Multivariate Statistics

[1]Hikmet Orhan, [2]Irfan Ozturk, [2]Zeki Dogan and [2]Sabri Yurtseven
[1]Department of Animal Science, Faculty of Agriculture, S. Demirel University, Isparta, Turkey
[2]Department of Animal Science, Faculty of Agriculture, Harran University Sanliurfa, Turkey

Abstract: In this study, livestock numbers and animal production of the main regions of Eastern and Southeastern of Turkey were evaluated by multivariate statistical analysis methods in order to determine the recent structure of livestock production. Data were taken from Turkish Statistical Institute (Turk Stat). The cities showing structural resemblance with each other were established as results of cluster analysis were summarized in dendograms. Also, accuracy of the clustering analysis was tested by Wilk's Lamda statistics value and assigning of cluster of cities was determined with discriminant analysis.

Key words: Multivariate statistics, discriminant, clustering, livestock production, East and Southeast Anatolia

INTRODUCTION

East and Southeast Anatolia regions have been poor derived regarding agricultural based industrial facilities because of both harsh geographic structure and climatic conditions. In these regions, agriculture and livestock based sectors, which are important for the region and has potential, did not yet able to show the expected.

Eastern Anatolian has about 39% of countries total pasture land and this is higher than countries average value (Anonymous, 2000). This number goes down to 11% in southeastern Anatolia (Sahin, 1999). Cities of Kars and Ardahan located in eastern Anatolia receive the highest amount of irrigation amounts resulting in high production of pasture. About 81.2% of farms in eastern Anatolia make both plant and animal production and only 7.6% of the total farms have livestock production and almost all of the farms are not modernized and specialized enterprises and the number of modern farming is extremely low in the region. Generally, farmers have 1-5 cows for household needs and animal based productions more than the need would be commercialized (Anonymous, 2000). In Eastern Anatolia, livestock production is practiced on the pastures in summers and in the completely closed burns in winter, but in tempered climate regions livestock is mostly practiced on the pasture in summer and in the semi-closed burns in winter (Anonymous, 2002-2004; Akbay and Boz, 2005; Anonymous, 2006). Present livestock status of these regions needs to be analyzed correctly for clarification of to why livestock farms in the region are insufficiency and under-developed.

In this study, distribution profile of the regions is taken out using multivariate statistical analyze methods in order to prevail the present status of animal breeding and production according to the cities of Eastern and South-Eastern Anatolia. Especially, similarity and non-similarity relations between cities and present status of the animals of these regions were to be shown with dendograms taken from cluster analyze.

MATERIALS AND METHODS

Data on number of animals and animal production levels were obtained from records of Turk Stat and GAP regional improvement managing presidency statistics. Data included average number of sheep and goat, cattle, poultry hire and production of meat, milk, egg, wool, leather and honeybee in 2002-2004 years (Anonymous, 2002-2004). For the easiness of the evaluation of data, sheep and goat numbers and cattle and buffalo numbers were combined. The poultry indicates only chicken and cock together while duck, Turkey, goose and other poultry numbers are interpreted as other poultry and for the sake of simplicity of evaluation poultry meat production were classified as broiler and laying hens. Egg production numbers in this study, indicates only chicken eggs. Sheep wool, goat hair and angora were evaluated in the same category.

Corresponding Author: Hikmet Orhan, Department of Animal Science, Faculty of Agriculture, S. Demirel University, Isparta, Turkey

Table 1: Multivariate data matrix

	Characters						
Provinces	1	2	3	.	.	.	p
1	X_{11}	X_{12}	X_{13}	.	.	.	X_{1P}
2	X_{21}	.	.	.	.	.	X_{2P}
3	X_{31}	.	.	.	.	.	X_{3P}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
n	X_{n1}	X_{n2}	X_{n3}	.	.	.	X_{np}

The data related to provinces are multivariable data and therefore those data were illustrated with X matrixes with n×p dimensions. Namely, n times provinces (in rows) would have p times characteristics (in column) as summarized in Table 1.

Or if, $X = (X_{ik})_{np}$; i = 1, 2,. .., n; k = 1, 2, ..., p

$$X = \begin{bmatrix} x_{11} & x_{12} & . & . & x_{1P} \\ x_{21} & . & . & . & x_{2P} \\ x_{31} & . & . & . & x_{3P} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{n1} & x_{n2} & . & . & x_{nP} \end{bmatrix}_{n\times p}$$

For a correct classification of provinces for animal production, provinces were classified with respect to animal numbers and animal production. To do this classification, the distance levels of provinces were determined. Even though, there were so many methods for calculation of the distance criteria, Euclidean distance given in Eq. 1 was the most suitable as the literature review (Edelbrock, 1979; Chirpaz *et al*., 2004; Sharma, 1996). The Euclidean distance Eq. 1 given in Eq. 3 is not affected by the unit of the studied variables. Therefore, Euclidean distance method was used for numeric classification. Furthermore, this method minimizes mean square error while, the ward clustering method uses the squared Euclidean distance. Main difference of ward clustering method from Euclidean distance methods is that this method does not take the square root of total distance according to variables into account. The results of this method can be taken by Eq. 2.

$$d(i,j) \Rightarrow d_2(x_i,x_j) = \left[\sum_{k=1}^{p}\left|x_{ik} - x_{jk}\right|^2\right]^{1/2} \quad (1)$$

$$d(i,j)^2 \Rightarrow d_2(x_i,x_j) = \sum_{k=1}^{p}\left|x_{ik} - x_{jk}\right|^2 \quad (2)$$

where:

i = 1, 2, …, n
j = 1, 2, …, n
k = 1, 2, …, p
n = The number of replicate
p = The variable numbers 9, 10

$$Z_{ik} = \frac{x_{ik} - \mu}{\sigma_k} \quad (3)$$

x_{jk} = Value of the kth variable for the jth entity
X_{ik} = Value of the kth variable for the ith entity
σ_k = Standard deviation of the kth variable

Ozturk *et al*. (1999) previously used ward method, which explained the superiority for classification of the cities.

Ward method is more general classification named as minimum variance method. The most important feature of this method is to give more reliable results than other classification methods when n>50 (Anderberg, 1973).

Statistically main goal of ward method is to minimize the variance within class. This method minimizes the class distribution. While, sometimes, there is danger of loosing data because of classification of entities in any step of classification. Losses can be estimated with the distance deviation square of the each point from its mean for every class. The second important characteristic of this method is constitution of similarity matrix elements using the squared Euclidean distance of Sum Square Error (SSE). Squared sum of Euclidean distance can be calculated with Eq. 4:

$$SSE_k = \sum_{i=1}^{p}\sum_{j=1}^{n_k} X_{ijk}^2 - \frac{(\sum_{i=1}^{p} X_{ik})^2}{n_k} \quad (4)$$

where:

X_{ijk} = ith score of jth individual of the k cluster

Total error sum of square within class can be calculated for k = 1, 2, …, n values using Eq. 5:

$$SSE = \sum_{i=1}^{k=n} SSE_k \quad (5)$$

After calculation of these values associated the classes can be done step by step.

First, assume SSE_k = 0 accepting every entity as a class. Second, step UV class is formed combining the U and V classes, which provides minimum increase. This can be calculated by Eq. 6:

$$\Delta SSE_{UV} = SSE_{(VU)} - SSE_{(U)} - SSE_{(V)} \quad (6)$$

In this way, n entities are split into n-1 classes.

Second step is repeated until cluster number is one (k = 1) and with this way individuals are linked step by step. Therefore, minimum increase in SSE with every step of the method is proportional to square of the Euclidean distance of mid-points of combined classes (Sharma, 1996; Su and Chou, 2001; Silva and Brito, 2006).

Wilk's lamda likelihood ratio test will used as test statistics. The method suggested from Wilk's is based on Generalized Likelihood Ratio Principle (GLRP). Test statistics (Λ) is estimated by Eq. 7.

$$\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|W+B|} \quad (7)$$

where:

B = Between groups
W = Within groups
T = Total source of variation in one way multivariate analyzes of variance (Ozdamar, 1999)

Moreover, test statistics take values from 0-1. When Λ is close to 0, it refuses H_0 hypothesis, while Λ is close to 1, it does not refuses H_0 hypothesis. Because of F test statistics is used in this method as critical value, Eq. 8 transformation is made:

$$F = \frac{1-\Lambda^{1/t}}{\Lambda^{1/t}} \cdot \frac{r \cdot t - 2 \cdot u}{p \cdot q} \sim F_{(p \cdot q,\, r \cdot t - 2 \cdot u)a} \quad (8)$$

In the Eq. 8, r = v - (p + q + 1)/2 and

$$t = \sqrt{\frac{(p^2q^2-4)}{(p^2q^2-5)}} \Rightarrow \text{for } (p^2q^2 - 5) > 0 \text{ then } t = 1$$

where:

u = (pq - 2)/4 q = k-1
n = Number of units in samples
p = Variable numbers
k = Group numbers
v = Error degrees of freedom (Byrne and Arnold, 1983; Ozdamar, 1999)

At the end of the cluster analyze, the discriminant analysis estimates whether, cities are located into right clusters or not. To do this, at the end of discriminant analysis a significant test of the estimated discriminant functions for separate power is tested with test statistics given in Eq. 9:

$$Q = \frac{(n - k\sum_{j=1}^{k} n_{ij})^2}{n(k-1)} \sim x^2_{(1;\alpha)} \quad (9)$$

where:

k = Number of groups
n = Number individual observations (Ozdamar, 1999)

MINITAB and STATISTICA statistical packages were used for analysis.

RESULTS AND DISCUSSION

Average animal data from 2002-2004 for the cities were classified as one by one according to animal numbers and animal production levels. Classification of the cities was given for animal numbers in Fig. 1 and for animal production Fig. 2.

Figure 1 indicates that similarity levels of the cities decreased from 100-0% cutting point. Dissimilarity levels among cities can be accepted as the point comes after zero cutting point of optimum cluster numbers. That is, optimum cluster numbers of cities to be separate is accepted as three.

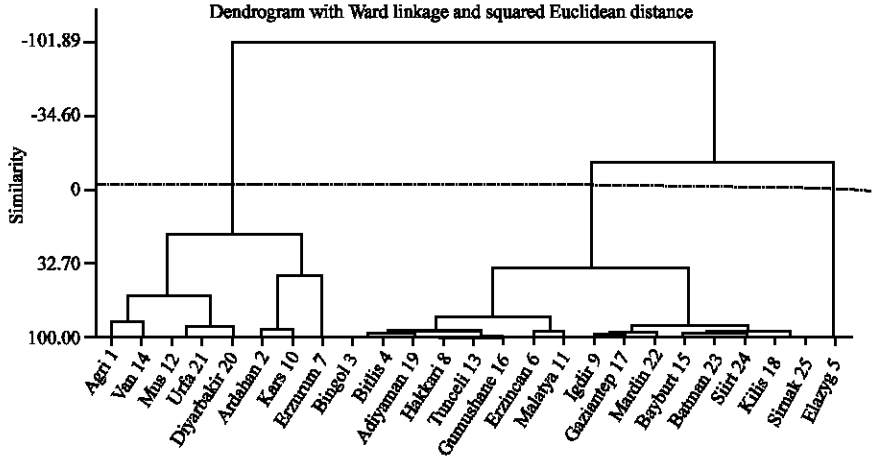

Fig. 1: Classification of the cities using Ward method according to number of animals

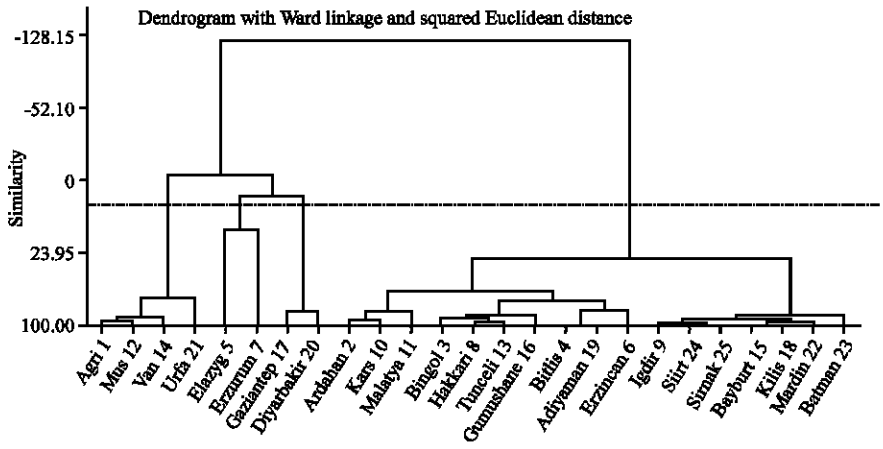

Fig. 2: Classification of the cities according to their animal productions

Table 2: Results of MANOVA Wilks' Lamda statistics

Characteristics	Test statistic	Approx F	p-value
Animal number	0.00321	59.988	0.000
Animal production values	0.00139	12.157	0.000
Including all the variables	0.00727	38.618	0.000

Table 3: Results of discriminant analyses

		City number (%) in every cluster				
Characteristics	Cluster number	1.00	2.00	3.00	4.00	Total
Animal production	1	4 (100.0)	0.0	0.0	0.0	4
	2	0.0	16 (94.1)	1 (5.9)	0.0	17
	3	0.0	0.0	2 (100.0)	0.0	2
	4	0.0	0.0	0.0	2 (100.0)	2
Animal numbers	1	8 (100.0)	0.0	0.0	-	8
	2	0.0	16 (100.0)	0.0	-	16
	3	0.0	0.0	1 (100.0)	-	1
Including all the variables	1	7 (100.0)	0.0	0.0	-	7
	2	0.0	17 (100.0)	0.0	-	17
	3	0.0	0.0	1 (100.0)	-	1

As shown in Fig. 2, cities were divided into 2 main clusters. But, below the cutting point for the dendogram cluster numbers were seen to be >2.

Multivariate Analysis of Variance (MANOVA) was used to test if levels of 3 groups of cities were separated significantly or not. Wilk's lambda value and F transformation value calculated by the Eq. 7 and 8, respectively and were given in the Table 2.

From Table 2, according to Wilks Lamda statistics there is a significantly difference among clusters ($p<0.001$).

Cluster analyze was used to see if the cities assigned to correct groups. Degree of accuracy estimating clustering results taken with discriminant analyses were given in Table 3. Analysis results indicated that calculations from Eq. 9 showed that there is 100% accuracy in assignation of cities in to the clusters.

To determine the variables affecting clustering of the cities according to their animal numbers were classified

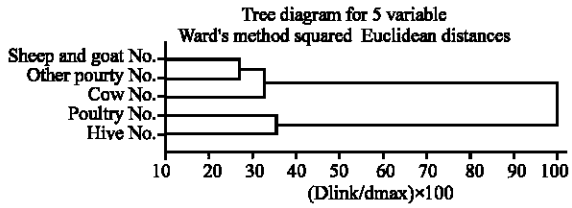

Fig. 3: Euclidian distances of the affecting variables for classification of the cities

with ward method and shown in dendongram (Fig. 3). It seemed that poultry and hive numbers affect the grouping of cities. The cities were significantly different from each other according to poultry and hive numbers.

It can be shown in Fig. 2 that contrary to previous grouping, cities of Elazig and Erzurum fall into the same group according to animal production value. Agri, Mus, Van and Sanliurfa all fall into the same group. The third group was combined the Gaziantep and Diyarbakir. These mentioned sub-groups were the same group in fundamental, but they were divided into sub-groups under the optimal cutting point. Lastly, the other all cities fall into the 4th group.

To test if the cities were assigned into the right clusters, discriminate analysis was performed and the results were given in Table 3. As shown in Table 3, all the cities assignment was 100% accurate according to animal number and including all the variables. Only one city was assigned to a wrong cluster for animal production.

MANOVA was used to test if levels of 3 groups of cities were separated significantly.

Right clustering ratio of the original groups can be calculated as:

$$\frac{\tilde{n}}{n} = \frac{24}{25} = 0.96$$

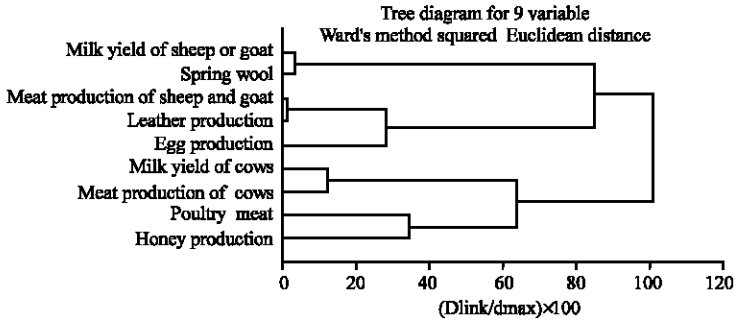

Fig. 4: Classifications of the animal production values

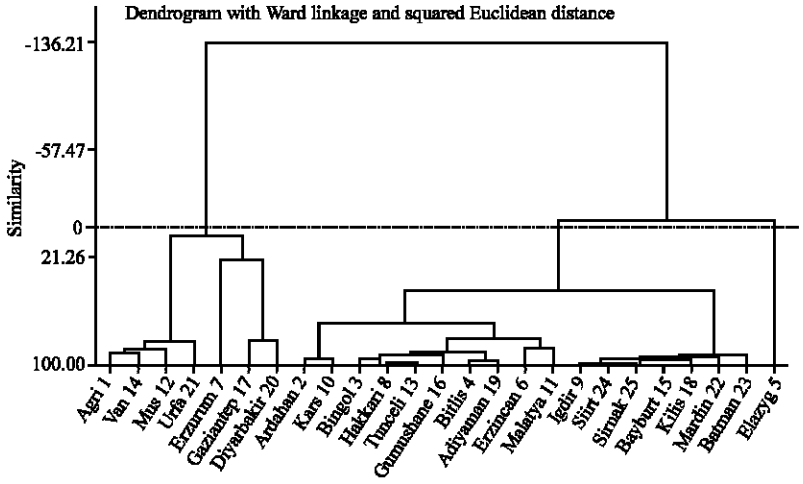

Fig. 5: Classifications of the cities according to animal numbers and animal production values

Where:

$$\tilde{n} = \sum_{j=1}^{k} n_{ij} = 4+16+2+2 = 24$$

Distribution results of the discriminant analysis were given in Table 3 according to Eq. 9 and calculated following result.

$$Q = \frac{(25-24\cdot 4)^2}{25\cdot(4-1)} = 67.21 > 6.63 = X^2_{1;\,0.01}$$

There were significantly difference among distribution clusters ($p<0.01$). This indicates that discriminant analysis separate the groups in the best way. Dendogram for classifications of the animal production were given in Fig. 4.

From the Fig. 4, the dendogram shows that cow milk and meat, poultry meat and honey productions have important role for separating the cities from each other.

When the cities were clustered based on both number of animals and animal productions, we obtained the dendogram shown in Fig. 5. There were significant difference among the clusters given in Fig. 5 ($p<0.001$). In that Fig. 5, there are 2 main groups. Under zero point groups closed to each other, so group numbers more than two. Cluster numbers were found to be 3 according to MANOVA and the Wilks, Lamda value was calculated as 0.00727 (Table 2).

Table 4: Classification of the studied cities according to their animal levels

Cluster	Levels	Cities
1	Rich	(Agri, Van, Mus, Urfa, Erzurum, Gaziantep, Diyarbakir)
2	Medium	(Elazig)
3	Poor	a-(Ardahan, Kars, Bingol, Hakkari, Tunceli, Gumushane, Bitlis, Adiyaman, Erzincan, Malatya)
		b-(Igdir, Siirt, Sirnak, Bayburt, Kilis, Mardin, Batman)

When the Fig. 5 examined deeply, the clustered cities can be summarized in Table 4. The variables clustered within them with ward method to determine which variables affected the clustering of cities results in Fig. 6 were obtained. Examination of the Fig. 6 in details for grouping the variables; it can be summarized in Table 5. Variables, which combined and separated the cities (Table 5).

In Table 4, first cluster formed the cities, which had big number of animals and animal production. Second

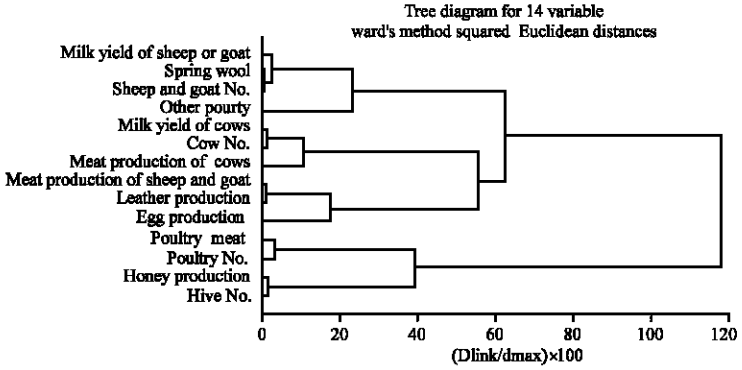

Fig. 6: Clusters of the variables that affect the grouping of the cities

Table 5: Classification the variables affecting the clustering of cities

Variables separated the cities	Variables combined the cities
	Milk yield of sheep or goat
Poultry meat	Spring wool
Poultry numbers	Sheep and goat numbers
Honey production	Other poutry
Hive numbers	Milk yield of cows
	Cow numbers
	Meat production of cows
	Meat production of sheep and goat
	Leather
	Egg production

cluster only had the city of Elazig. Elazig is medium city according to animal numbers and animal production. There was important role of poultry numbers, poultry meat, honey production and hive numbers to form only group Elazig. Because, Elazig city was the first queue in accordance with poultry numbers and poultry productions in 2002-2004. Also, it was the first three queue in accordance with honeybee breeding. In third cluster, called as poor separated 2 groups. First group was richer than second. It is concluded that if we want to increase the regions in accordance with animal production should be pay attention Table 5.

CONCLUSION

Over all, in case of the number of animals and animal production, cities were shown approximately similar characteristics. And also, in all dendograms, there seemed to be 2 main clusters at the beginning and than in advanced stages formed 3 or 4 clusters. According to Fig. 5, in which all the variables evaluated together, there was significant difference in 3 groups.

REFERENCES

Akbay, C. and I. Boz, 2005. Turkey's livestock sector: Production, consumption and policies. Livestock Research for Rural Development, 17 (9). http://cipav.org.co/lrrd/lrrd17/9/akba17105.htm.

Anderberg, M.R., 1973. Cluster Analysis for Applications. 1st Edn. Academic Press, New York, pp: 10-24. ISBN: 0-12-057650-3. http://adsabs.harvard.edu/abs/1973caa..book.....A.

Anonymous, 2000. Agricultural main report of East Anatolia (Present status and analyze). Vol. II, DPT, Ankara-Turkey. http://ekutup.dpt.gov.tr/bolgesel/dap/durum2.pdf.

Anonymous, 2002-2004. Turk Stat, Agricultural structure (Production, cost, value). Turkish Statistical Institute, Ankara. http://www.tuik.gov.tr/VeriBilgi.do?tb_id=46&ust_id=13.

Anonymous, 2006. Common market coordination working group reports for meat of Cow, bull, sheep, goat and milk and milk production, egg and poultry, rough feed, silkworm breeding, honey and beekeeping. T.C. Ministry of Agriculture and Rural Affairs-Strategy Development, Ankara. http://sgb.tarim.gov.tr/yayimlar/cilt_1.pdf.

Byrne, P.J. and S.F. Arnold, 1983. Inference About Multivariate Means for a Nonstationary Autoregressive Model. J. Am. Stat. Assoc., 78 (384): 850-855. http://www.jstor.org/stable/2288195.

Chirpaz, E., M. Colonna and J.F. Viel, 2004. Cluster analysis in geographical epidemiology: The use of several statistical methods and comparison of their results. Epidemiol Sante Publique, 52 (2): 139-149. PMID: 15138393. http://www.cababstractsplus.org/google/abstract.asp?AcNo=20043085514.

Edelbrock, C., 1979. Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody. Multivariate Behavioral Research. DOI: 10.1207/s15327906mbr1403_6. http://web.ebscohost.com/ehost/detail?vid=1&hid=108&sid=b1d5be5f-50ae-4b4b-a4b6-2f1bfe47ebe4%40sessionmgr102&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#db=pbh&AN=6378322.

Ozturk, I., Z. Dogan, R. Sirali and N. Yildiz, 1999. Comparison of honeybee genotypes belong to different regions of Turkey with single linkage clustering, complete linkage clustering and ward methods. GAP 1. Agricultural Congress, pp: 26-28. Sanliurfa, Turkey.

Ozdamar, K., 1999. Data analyze with statistical packaged software-2 (MANOVA). 2nd Edn. Paket programlar ile istatistiksel veri analizi-2 (Çok Degiskenli Analiz), Kaan Kitabevi, Eskisehir-Turkey, pp: 137-183. ISBN: 9756787015.

Su, M.C. and C.H. Chou, 2001. A modified version of the k-means algorithm with a distance based on cluster symmetry. IEEE. Trans. Pattern Analysis and Machine Intelligence, 23 (6): 674. DOI: 10.1109/34.927466.

Sharma, S., 1996. Applied Multivariate Techniques. 1st Edn. John Wiley and Sons. Inc. New York, pp: 185-195. ISBN: 978-0-471-31064-8. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471310646,descCd-tableOfContents.html.

Silva, A.P. and P. Brito, 2006. Linear discriminant analysis for interval data. Computational Statist., 21: 289-308. DOI: 10.1007/s00180-006-0264-9. http://portal.acm.org/citation.cfm?id=1182392.1182401&coll=&dl=acm.

Sahin, K., 1999. Structural and economical problems negatively effect on live stock breeding in East Anatolia region and solution and suggestion to the problems. Int. Anim. Congress, Izmir-Turkey, pp: 590-595. http://www.yyu.edu.tr/calismaayrinti.aspx?sicilno=1015&adsoyad=Ara%C5%9F.G%C3%B6r.Dr.%20%20%20Kas%C4%B1m%20%20%20%C5%9EAH%C4%B0N.